



Improving Data Collection Strategies

JANE L. DELGADO, PhD
LEO ESTRADA, PhD

Dr. Delgado is President and Chief Executive Officer of the National Coalition of Hispanic Health and Human Services Organizations, also known as COSSMHO. Dr. Estrada is a Professor at the University of California, Los Angeles, at the Graduate School of Architecture and Urban Planning.

Tearsheet requests to Jane Delgado, PhD; President, National Coalition of Hispanic Health and Human Services Organizations, 1501 16th St., NW, Washington, DC 20036; tel. (202) 797-4321; fax (202) 797-4353.

Synopsis

The authors examined 21 major health data systems of the U.S. Department of Health and Human Services (DHHS) and concluded that data on Hispanics are not included in several Departmental national health data collection systems, and that even when collected, data on Hispanic subpop-

ulations are found in few of the systems. Of the 21 data systems, 6 do not collect Hispanic population data, including the Medicare statistical system. Only the National Vital Statistics System was found to collect data for all major Hispanic subpopulation groups. Seventeen of the 21 data systems do not collect sample sizes adequate for analyzing any one of the four major Hispanic subpopulation groups.

To address that lack, Hispanic health leadership agendas have recommended collecting data on Hispanics in all systems, where possible, to provide samples of a size adequate for detailed analysis of Hispanic subpopulation groups, for support of researchers and a Hispanic research infrastructure, and for broad dissemination of data findings, including dissemination in useful formats to Hispanic community-based organizations.

IN 1970, THE DECENNIAL CENSUS called for self-identification of “persons of Spanish origin,” as it was termed, thereby including all Hispanics in a national data system for the first time (1). Prior to 1970, Hispanics were identified in the census by means of such proxies as Spanish surname, Spanish speaking, or birthplace and birthplace of parents.

In 1976, health data collection moved onto the national agenda, when Congress enacted Public Law 94-311, which called on “the Department of Health, Education, and Welfare (to) collect, and publish regularly, statistics which indicate the * * * health * * * condition of Americans of Spanish origin or descent” (2). In 1978, the Office of Management and Budget released Directive 15, “Race and Ethnic Standards for Federal Statistics and Administrative Reporting,” which set the standard for the categorization of Hispanic and non-

Hispanic persons under national data systems (3). However, the health component of Public Law 94-311 was not implemented.

In 1977, the National Academy of Public Administration was contracted to evaluate the National Health and Nutrition Examination Survey (NHANES), which is conducted by the Centers for Disease Control and Prevention through its National Center for Health Statistics (NCHS). The Academy recommended that NCHS consider conducting studies of the health and nutrition status of major subpopulations, studies that could yield estimates comparable to those for the general population (4). As a result, from 1982 to 1984, NCHS conducted the Hispanic Health and Nutrition Examination Survey (HHANES), which was the first special population survey undertaken by NCHS and the first comprehensive Hispanic health

survey to be carried out in the United States (4). However, HHANES represented a one-time-only study.

In 1985, the Secretary of the Department of Health and Human Services (DHHS), Margaret Heckler, released the landmark "Report of the Secretary's Task Force on Black and Minority Health." The report documented, for the first time, the areas in which health data comparable to those reported for other racial groups were unavailable for Hispanics, such as life expectancy, congenital defects, and morbidity trends. Because only limited Hispanic health data were available, the task force commissioned special studies, used regional data, and conducted an extensive review of Hispanic health literature in search of available data to report (5).

The Secretary's report provided the impetus for including a Hispanic identifier on the national model birth and death certificates in 1989. As a result, Hispanic birth and mortality data are currently available from 44 States and the District of Columbia, representing 97 percent coverage of the Hispanic population (6). Although significant progress has been made in the area of Hispanic mortality data reporting, significant gaps remain in the collection and reporting of Hispanic morbidity data.

In response to the need for essential Hispanic health data, Congress in 1990 passed Section 7 of the Disadvantaged Minority Health Improvement Act (7). The provision called on NCHS to "collect and analyze adequate health data that is specific to particular ethnic and racial populations, including data collected under national health surveys." The congressional mandate has promoted the inclusion of statistically valid Hispanic samples in some current national health data systems. However, sampling by Hispanic subgroups remains limited.

In recent years, the Current Population Survey produced by the Bureau of the Census has provided data on health insurance coverage for a sample in which Hispanics have been oversampled. The sample includes about 20,000 Hispanic Americans, providing data on health insurance coverage of Hispanics as a group and by national origin.

The 21 major DHHS health data systems reported in "Health, United States, 1990" (8), are shown in the table. Those data systems largely form the basis of our quantitative knowledge of public health in the United States. However, analysis of the 21 data systems reveals two major characteristics: Hispanics are not included in several DHHS national health data collection systems,

and little data specific for Hispanic subpopulations are available. Of the 21 data systems, 6 do not collect Hispanic population data, including the Medicare statistical system. Furthermore, of these data systems, only the National Vital Statistics System collects data for all Hispanic subpopulation groups. Seventeen of the 21 data systems do not collect samples of a size adequate for analysis for any of the four major Hispanic subpopulation groups.

Of the 21 data systems, 3 provide the majority of population morbidity data: NHANES, the National Health Interview Survey (NHIS), and the National Medical Expenditures Survey (NMES). Each of those systems selects a representative population sample that is used to arrive at nationally representative morbidity and health utilization figures. NHANES oversamples Mexican Americans, NHIS oversamples Hispanics, and NMES oversamples Hispanics and Mexican Americans. The sampling frameworks of the national samples are not designed to produce valid figures for Hispanic populations (9). To arrive at representative morbidity figures for them, it is necessary to redesign the samples to oversample from Hispanic population groups or to merge data for several years to obtain adequate sample sizes.

Oversampling for Hispanic subpopulations is urgent, because new data show that there are significant morbidity differences among those subpopulation groups, such as, for example, the prevalence of hypertension. At one time, geographic areas were used as proxies for subpopulation groups: New York represented Puerto Ricans, Florida represented Cuban Americans, and Texas represented Mexican Americans. However, demographic shifts have invalidated those proxies. Providing Hispanic oversamples and sufficient subpopulation detail for health planning and delivery is a critical data collection issue for Hispanic communities (10).

An immediate effect of the lack of Hispanic mortality, morbidity, and health services utilization data has been the lack of Hispanic-specific objectives under the Healthy People 2000 (11) initiative. Healthy People 2000 is a DHHS effort to detail and set the nation's disease prevention and health promotion goals. Because of a lack of baseline data to assess risk, only 25 of 300 Healthy People 2000 objectives contain specific components focusing on the Hispanic community. As a result, there are no Hispanic component objectives in the areas of alcohol and other drugs, mental health and mental disorders, unintentional injury, occupational safety and health, environmental health, food and drug

Twenty-one major data collection systems of the Department of Health and Human Services, reported in "Health, United States, 1990" (8), showing Hispanic and Hispanic subgroup data that are collected in sample sizes adequate for analysis, by agency

Agency and system	Hispanic	Mexican American	Puerto Rican	Cuban American	Central-South American
Agency for Health Care Policy and Research:					
National Medical Expenditures Survey (24) ¹	X	X
Centers for Disease Control and Prevention:					
National Center for Infectious Diseases:					
AIDS Surveillance (25)	X
National Center for Prevention Services:					
U.S. Immunization Survey (26) ²	X
Epidemiology Program Office:					
National Notifiable Diseases Surveillance System (27)	X
National Center for Health Statistics (NCHS):					
National Ambulatory Medical Care Survey (28a) ³	X
National Health and Nutrition Examination Survey (28b, 29)		X
National Health Interview Survey (28c) ⁴	X
National Hospital Discharge Survey (30)	X
National Maternal and Infant Health Survey (31)	⁵ X
National Survey of Family Growth (32) ⁶	X
National Vital Statistics System (5, 33)	X	X	X	X	X
National Institute for Occupational Safety and Health:					
National Occupational Hazard Survey (34)
National Occupational Exposure Survey (35)
Health Care Financing Administration (HCFA):					
Medicaid Data System (36, 37)	X
Medicare Statistical System (38) ⁷
Health Resources and Services Administration:					
Physician Supply Projections (39) ⁸	X
Nurse Supply Estimates (39) ⁹	X
National Institutes of Health:					
National Cancer Institute:					
Surveillance, Epidemiology, and End Results Program (40) ¹⁰	¹¹ X
National Institute on Alcohol Abuse and Alcoholism:					
National Survey of Drinking (41)
National Institute on Drug Abuse:					
National Household Survey on Drug Abuse (42)	X
Drug Abuse Warning Network (43)	X

¹ Unweighted number of Mexican American respondents is 2,590. 100 respondents are generally considered the minimum number of respondents for statistical validity. While there is an oversample for Hispanics, but not one for Mexican Americans, there are enough Mexican American respondents to arrive at statistically valid figures for this subpopulation for most items under the survey, except when broken down by some geographic indicators.

² Information was obtained also in personal communication from Ann Hardy, NCHS, Division of Health Interview Statistics. The 1991 and 1992 Health Interview Survey included immunization coverage questions. The 1992 sample was expected to yield estimates for Hispanic and Mexican American children younger than 6 years, although it is unclear whether estimates will be available for 1-year age groupings.

³ Information was obtained also in personal communication from Raymond Gagnon, NCHS, Ambulatory Care Statistics Branch. The number of Hispanic-identified patient visits in the 1990 fielding of the survey was 35,456,000, with 522,000 visits being the minimum number of visits required for a statistically valid sample.

⁴ Information was obtained also in personal communication from Owen Thornberry, NCHS, Division of Health Interview Statistics. The 1991 and 1992 Health Interview Surveys were combined to create a Hispanic oversample.

⁵ Texas only.

⁶ Information was obtained also in personal communication from William Pratt, NCHS, National Survey of Family Growth. Although Hispanics were not oversampled under the National Survey of Family Growth, in Cycle IV, out of 8,450 women sampled, 641 were Hispanic, allowing for statistically valid data runs on a majority of questions in the survey with regard to Hispanic women. Cycle V, to be fielded in 1994, is expected to oversample for Hispanics.

⁷ Information was obtained also in personal communication from Dick Lyman, HCFA.

⁸ Information was obtained also in personal communication from Gloria Bronstein, HCFA, Bureau of Health Professions, Office of Data Analysis and Management. While physician supply projection data were not available by subgroups, data on Mexican Americans and mainland Puerto Ricans were available for numbers of current medical school students and graduates.

⁹ Information was obtained also in personal communication from Gloria Bronstein, HCFA, Bureau of Health Professions, Office of Data Analysis and Management.

¹⁰ Data were available for the Commonwealth of Puerto Rico without ethnic classification.

¹¹ New Mexico only.

safety, and sexually transmitted diseases. There are currently no plans to supplement the number of Hispanic component objectives under Healthy People 2000.

In addition to collection of adequate data on Hispanic subpopulations, timely data analysis is necessary to ensure that the information becomes available to practitioners through the public health literature and is quickly put to use in the planning

and delivery of health care services to Hispanic communities. To accomplish that goal, in 1986 the National Coalition of Hispanic Health and Human Services Organizations, also known as COSSMHO, established the Hispanic Health Research Consortium (HHRC) with funding from NCHS. HHRC serves as a university-based Hispanic health research infrastructure that supports timely analysis of data from HHANES and other health surveys

containing adequate sample sizes for detailed analyses of Hispanic populations. Three rounds of competitive grants were awarded by COSSMHO, the latest in 1992, to establish a network of 13 research teams throughout the country.

Research from the first two rounds of HHRC programs was published in December 1990 as an 11-article supplement to the American Journal of Public Health, "Hispanic Health and Nutrition Examination Survey, 1982-84: Findings on Health Status and Health Care Needs (12)." The following month, the Journal of the American Medical Association published its "Hispanic Health Issue" (13), which included papers from an HHRC team. Those publications are benchmark contributions to the health care literature and demonstrate the need to support timely analyses and dissemination of the results of Hispanic health research.

Recommendations

During recent years, Hispanic leadership has developed agendas for health. The authors reviewed them and related documents to determine community priorities for Hispanic health data collection, analysis, and dissemination (12-21). The following is a summary of major Hispanic leadership consensus recommendations to DHHS that continue to require attention.

Data collection.

1. Include in all major DHHS health data collection systems statistically valid samples representative of Hispanic populations and subpopulations.
2. Redesign morbidity and health utilization surveys to obtain critical subgroup information to guide practitioners and policy makers.
3. Redesign samples to collect data with more population specificity in the *Central-South American* and the *Other Hispanic* subpopulation categories, particularly for data specific to the growing Dominican and Salvadoran community.
4. Establish Hispanic component objectives for each Healthy People 2000 objective and the data collection systems necessary to track progress toward those component objectives.
5. Assess the validity of current data collection instruments and procedures for data collection used in Hispanic communities.

Data analysis.

1. Support the development of a Hispanic health

research infrastructure to analyze Hispanic information from existing health data collection efforts.

2. Provide technical assistance to Hispanic community-based organizations and Hispanic researchers applying for research grants for data analysis awards.

3. Include rating criteria under data analysis requests for applications (RFA) and requests for proposals (RFP) that require the applicant to demonstrate cultural competency before conducting any proposed study.

4. Include Hispanic researchers in the development of RFAs and RFPs for data analysis and in peer review committees to assess submitted proposals.

5. Submit an annual report of relevant activities conducted and expenditures made to improve Hispanic health statistics, including the percentage of data analysis grant funding awarded to programs focused on Hispanic populations.

Data dissemination.

1. Support the dissemination of Hispanic health research findings to lay audiences, including the press, policymakers, and community-based organizations.

2. Continue to include data on Hispanic subpopulations in annual summary reports made by Departmental and Public Health Service agencies, particularly the "Health, United States," series.

3. Make data more accessible to health researchers and community-based organizations without access to mainframe computers.

Considerations for Implementation

Data collection. If the Department is to enhance inclusion of data on Hispanic populations in its data collection systems, issues of standardizing data collection need to be addressed. In particular, the Public Health Service should publicize and reissue the current Standards for the Collection and Analysis of Racial and Ethnic Data in Federal Agencies, Office of Management and Budget Statistical Policy Directive 15 (22), as called for by the Public Health Service Task Force on Minority Health Data. The standards outline a minimum set of race and ethnic categories and definitions required for use in all Federal data collection activities that include race items. PHS should develop and encourage the use of a supplemental set of standards for classifying racial and ethnic subpopulations for use in data activities. Given the growing diversity of

'This effort will require DHHS to reach out to Hispanic researchers, support the development of a Hispanic research infrastructure, and work with new community research partners, including Hispanic community-based health and human services organizations.'

the Hispanic population and the flexibility of the directive, it is now necessary and possible to update that directive to include standards for classification of major Hispanic subpopulations (23).

Furthermore, in addition to refining classifications of Hispanic subpopulations, an assessment of data collection methods and instruments is necessary to determine their adequacy for Hispanic populations. The instruments and data collection standards developed for HHANES are logical starting points. Other health agencies should be integrated into the process, since consensus must be attained. Finally, each departmental survey should be assessed to determine what savings could be realized from reducing the number of questions or clinical tests implemented under the survey. Any savings may be applied to expand the scope of the survey to include an oversample for small Hispanic populations.

Data analysis. The Department must enrich the relationships it has developed in the analysis of Hispanic health data. Hispanic researchers have not been encouraged to be a significant part of the research community. DHHS must therefore develop innovative approaches to support current Hispanic health researchers and foster the next generation of Hispanic research talent.

This effort will require DHHS to reach out to Hispanic researchers, support the development of a Hispanic research infrastructure, and work with new community research partners, including Hispanic community-based health and human services organizations. DHHS must set concrete goals for the support of Hispanic health data analysis and the inclusion of Hispanic health researchers.

Data dissemination. A final area of implementation to consider is the use of a morbidity model, instead of a mortality model, in reporting health data. Although DHHS has improved its data base efforts

related to Hispanic mortality, death rates tell a small part of the Hispanic health story. To plan health services and meet the health needs of Hispanic communities, researchers must have access to morbidity data to understand what occurs between birth and death. Reporting morbidity data must become a priority for those concerned with Hispanic health; DHHS must therefore prioritize the dissemination of Hispanic morbidity data in its health data reporting efforts. With the challenge of the collection of Hispanic mortality data largely met, the remaining challenge for the Hispanic leadership as well as for the health research community is to meet the need for data on morbidity in Hispanic communities.

References

1. Bureau of the Census: Persons of Spanish surname, 1970 Census of Population. PC(2)-ID. U.S. Department of Commerce. U.S. Government Printing Office, Washington, DC, 1973, p. x.
2. 94th Congress: H.J. Res. 92, Joint resolution relating to the publication of economic and social statistics for Americans of Spanish origin or descent (Public Law 94-311). U.S. Government Printing Office, Washington, DC, June 16, 1976.
3. Hahn, R.: The state of federal health statistics on racial and ethnic groups. JAMA 267: 268-271, Jan. 8, 1992.
4. Delgado, J. L., Johnson, C. L., Roy, I., and Trevino, F. M.: Hispanic Health and Nutrition Examination Survey, 1982-84: findings on health status and health care needs. I. Hispanic Health and Nutrition Examination Survey: methodological considerations. Am J Public Health 80 (Supp.): 6-10 (1990).
5. U.S. Department of Health and Human Services: Hispanic health issues. In Report of the Secretary's Task Force on Black and Minority Health, Vol. VIII. U.S. Government Printing Office, Washington, DC, 1985, p. 53.
6. National Center for Health Statistics: Advance Report of Final Mortality Statistics, 1989. Monthly Vital Statistics Report, Vol. 40, No. 8, Supp. 2. Centers for Disease Control, Hyattsville, MD, Jan. 7, 1992.
7. 101st Congress: Disadvantaged Minority Health Improvement Act (Public Law 101-527). U.S. Government Printing Office, Washington, DC, 1990.
8. National Center for Health Statistics: Health, United States, 1990. DHHS Publication No. (PHS) 91-1232. Centers for Disease Control, Hyattsville, MD, 1991.
9. U.S. Government Accounting Office: Health and nutrition: collection of vital and statistical data on Hispanics. U.S. Government Printing Office, Washington, DC, March 1988, p. 8.
10. Isaacs, M.: The complexity of minority health: an overview of key issues. Paper presented at the Conference on Setting a Research Agenda: Challenges for Minority Health Grants, Dec. 4-6, 1991. National Center for Health Statistics, Hyattsville, MD.
11. Public Health Service: Healthy people 2000: national health promotion and disease prevention objectives. DHHS

- Publication No. (PHS) 91-50212. Office of the Assistant Secretary for Health, Office of Disease Prevention and Health Promotion. U.S. Government Printing Office, Washington, DC, 1990.
12. Hispanic Health and Nutrition Examination Survey, 1982-84: findings on health status and health care needs. *Am J Public Health* 80 (supp.): 1-72, Dec. 1990.
 13. Hispanic Health Issue. *JAMA* 265: 161-296, Jan. 9, 1991.
 14. U.S. Government Accounting Office. Hispanic access to health care: significant gaps exist. GAO/PMED-92-6. U.S. Government Printing Office, Washington, DC, January 1992.
 15. Health Resources and Services Administration: Health care for Hispanic individuals. Report to Congress. DHHS Publication No. (HRSA) 240-89-0045. U.S. Government Printing Office, Washington, DC, 1991.
 16. National Coalition of Hispanic Health and Human Services Organizations: And access for all: Medicaid and Hispanics. Washington, DC, 1990.
 17. National Coalition of Hispanic Health and Human Services Organizations: Healthy People 2000: an ounce of prevention. Washington, DC, 1992.
 18. National Coalition of Hispanic Health and Human Services Organizations: The state of Hispanic health. Washington, DC, 1992.
 19. National Council of La Raza: Hispanics and health insurance status, Vol. 1. Washington, DC, 1992.
 20. National Council of La Raza: The state of Hispanic America, 1991: an overview. Washington, DC, 1992.
 21. U.S. Department of Health and Human Services: National Hispanic Health Policy Summit report. Office of the Assistant Secretary for Health, Office of Minority Health. U.S. Government Printing Office, Washington, DC, 1992.
 22. U.S. Office of Management and Budget: Statistical policy directive 15. U.S. Government Printing Office, Washington, DC, 1978.
 23. Hahn, R.: The state of federal health statistics on racial and ethnic groups. *JAMA* 267: 268-271, Jan. 8, 1992.
 24. Cohen, S., DiGaetano, R., and Waksberg, J.: Sample design of the 1987 household survey. National Medical Expenditures Survey, Methods 3. AHCPH Publications No. 91-0037. Public Health Service, Agency for Health Care Policy and Research, Rockville, MD, Sept. 1991.
 25. Centers for Disease Control: HIV/AIDS surveillance report. Atlanta, GA, July 1992, pp. 1-18.
 26. Adams, D.: Memorandum on release of the CPS Immunization/Smoking Public Use File. Current Population Survey, U.S. Department of Commerce, Washington, DC, 1992.
 27. Centers for Disease Control. Summary of notifiable diseases, United States, 1990. *MMWR Morb Mortal Wkly Rep* 39: 1-12, Oct. 4, 1991.
 28. Powell, A., and Golden, P.: Obtaining data from the National Center for Health Statistics on the health of minorities. Centers for Disease Control and Prevention, National Center for Health Statistics, Hyattsville, MD, 1989, (a) pp. 23-24; (b) 18; (c) pp. 12-13.
 29. National Center For Health Statistics: Summary of surveys and data systems. Centers for Disease Control, Hyattsville, MD, April 1992.
 30. Graves, E. J.: Detailed diagnoses and procedures, National Hospital Discharge Survey, 1988. National Center for Health Statistics. *Vital Health Stat* 10 [107] 133 (1991).
 31. Gonzalez, J. F., et al.: Sample design for the 1988 National Maternal and Infant Health Survey. National Center for Health Statistics, Hyattsville, MD, 1991, p. 1.
 32. Judkins, D. R., Mosher, W. D., and Botman, S.: National Survey of Family Growth: design, estimation, and inference. DHHS Publication No. (PHS) 91-1386. National Center for Health Statistics. *Vital Health Stat* 2 [109] 1991.
 33. Maurer, J. D., Rosenberg, H. M., and Keemer, J. B.: Deaths of Hispanic origin, 15 reporting states, 1979-81. DHHS Publication No. (PHS) 91-1855. Centers for Disease Control, National Center for Health Statistics. *Vital Health Stat* 20 [18] 1991.
 34. National Institute for Occupational Safety and Health: National Occupational Hazard Survey. Vol. I: Survey manual. NIOSH Publication 74-127. Centers for Disease Control, Cincinnati, OH, 1974.
 35. National Institute for Occupational Safety and Health: National Occupational Hazard Survey, field guidelines. DHHS Publication No. (NIOSH) 86-116. Centers for Disease Control, Cincinnati, OH, 1986.
 36. Health Care Financing Administration: Form HCFA-2082, Statistical report on medical care: eligibles, recipients, payments, and services. U.S. Government Printing Office, Washington, DC, 1991.
 37. Howe, C., and Terrell, R.: Health care financing program statistics. HCFA Publication 03249. Health Care Financing Administration, U.S. Government Printing Office, Washington, DC, 1987.
 38. Health Care Financing Administration: Medicare statistical file manual. HCFA Publication 03272. U.S. Government Printing Office, Washington, DC, 1988.
 39. Preliminary data from the Eighth Report to the President and Congress on the Status of Health Personnel in the United States. Health Resources and Services Administration, Rockville, MD, 1993.
 40. Sondik, E.: Cancer statistics review. National Institutes of Health Publication 89-2789. National Cancer Institute, Bethesda, MD, 1989.
 41. Reed, P. G., and Sanchez, D. S.: National Alcoholism and Drug Abuse Program Inventory. National Institute on Alcohol Abuse and Alcoholism, Division of Biometry and Epidemiology, Rockville, MD, 1986.
 42. National Institute on Drug Abuse: National Household Survey on Drug Abuse, Population Estimates 1991. Division of Epidemiology and Prevention Research. Rockville, MD, 1991.
 43. Colliver, J. D.: Characteristics and implications of the new DAWN emergency room sample. CEWG Report. 414: National Institute on Drug Abuse, Rockville, MD, June 1991.